Normative Data for the Words-in-Noise Test for 6- to 12-Year-Old Children

Richard H. Wilson
James H. Quillen VA Medical Center, Mountain Home, TN, and East Tennessee State University, Johnson City

Nicole M. Farmer
Avni Gandhi
Emily Shelburne
Jamie Weaver
East Tennessee State University

Purpose: To establish normative data for children on the Words-in-Noise Test (WIN; R. H. Wilson, 2003; R. H. Wilson & R. McArdle, 2007).
Method: Forty-two children in each of 7 age groups, ranging in age from 6 to 12 years (n = 294), and 24 young adults (age range: 18–27 years) with normal hearing for pure tones participated. All listeners were screened at 15 dB HL (American National Standards Institute, 2004) with the octave interval between 500 and 4000 Hz. Randomizations of WIN Lists 1, 2, and 1 or WIN Lists 2, 1, and 2 were presented with the noise fixed at 70 dB SPL, followed by presentation at 90 dB SPL of the 70 Northwestern University Auditory Test No. 6 (T. W. Tillman & R. Carhart, 1966) words used in the WIN. Finally, the Peabody Picture Vocabulary Test—Revised (L. M. Dunn & L. M. Dunn, 1981) was administered. Testing was conducted in a quiet room.
Results: There were 3 main findings: (a) The biggest change in recognition performance occurred between the ages of 6 and 7 years; (b) from 9 to 12 years, recognition performance was stable; and (c) performance by young adults (18–27 years) was slightly better (1–2 dB) than performance by the older children.
Conclusion: The WIN can be used with children as young as 6 years of age; however, age-specific ranges of normal recognition performance must be used.

KEY WORDS: auditory perception, children, hearing loss, speech perception, speech recognition in multitalker babble

The Words-in-Noise Test (WIN) was developed to quantify the ability of adults to understand speech in a background of multitalker babble (Wilson, 2003; Wilson & McArdle, 2007). It uses the Northwestern University Auditory Test No. 6 (NU No. 6; Tillman & Carhart, 1966) monosyllabic words presented at multiple signal-to-noise ratios (SNRs) to generate a psychometric function from which the 50% correct point can be calculated with the Spearman–Kärber equation (Finney, 1952). The 50% point provides a measure of hearing loss with respect to SNR (i.e., SNR hearing loss). The WIN was originally developed with 10 unique words presented at each of seven SNRs from 24- to 0-dB S/N in 4-dB decrements. For clinic implementation, the 70-word version was subsequently divided into two 35-word lists in which five words were presented at each SNR (Wilson & Burks, 2005). A 50% correct point at ≤6-dB S/N was established as the upper range of normal performance, which was defined by the 90th percentile from a group of 24 young adults with normal hearing for pure tones (Wilson, Abrams, & Pillion, 2003). In multiple studies, the WIN has demonstrated consistently the ability to separate recognition performances by adults with normal hearing for pure tones from performances by adults with pure-tone hearing loss (e.g., Wilson, Carnell, & Cleghorn, 2007; Wilson, McArdle, & Smith, 2007). The purpose of the present study was to extend the WIN norms to children in the 6- to 12-year age range.

Although the peripheral auditory mechanism appears to mature early in life, the neural processing that accompanies auditory function has a relatively protracted development period (Dawes & Bishop, 2008). A comparison of speech recognition/identification performances on a given set of materials by preschool and school-age children with normal hearing and by young adults with normal hearing for pure tones exemplifies this issue by providing two major differences between the two listener groups (Eisenberg, Shannon, Martinez, Wygonski, & Boothroyd, 2000; Elliott et al., 1979; Fallon, Trehub, & Schneider, 2000; Goldman, Fristoe, & Woodcock, 1970; Jamieson, Kranjc, Yu, & Hodgetts, 2004; Keogh, Kei, Driscoll, & Khan, 2010; McCreery et al., 2010; Nittrouer & Boothroyd, 1990; Palva & Jokinen, 1975; Stuart, 2005; Stuart, Givens, Walker, & Elangovan, 2006). First, in the quiet listening condition, and even on an identification task,[1] performances by children are poorer than performances by adults. Second, when the listening task is degraded in some manner (e.g., with the addition of background noise), performance differences observed in quiet between children and young adult listeners become exaggerated.

Siegenthaler's (1969, 1975) reports exemplify the work that has been conducted with children in the quiet listening domain. Siegenthaler used the Threshold by Identification of Pictures (TIP; Siegenthaler, Pearson, & Lezak, 1954) and the Discrimination by Identification of Pictures (DIP; Siegenthaler & Haspiel, 1966) tests to establish speech recognition thresholds (SRTs) and speech identification performances of 295 children 3 to 8 years of age. Recorded materials were used in a sound-field environment. The SRT with the TIP improved linearly from 17.9 dB SPL (3 years) to 9.4 dB SPL (8 years) with a slope of approximately −1.8 dB/year. Identification performance on the DIP, which was measured at the level of the SRT and at 5 and 10 dB above the SRT, improved overall from 50.3% (3 years) to 68.3% (6 years), which provided a slope of 6% per year. Performance above 6 years of age was asymptotic. Similarly, using the Northwestern University Children's Perception of Speech materials (NU-CHIPS), Elliott and Katz (1980) reported psychometric functions for 3-, 5-, and 10-year-old children and for adults. At the 60% point on the psychometric functions (their Figure 1), the performances by the 10-year-old children and adults were the same (2 dB SL re: the mean of the two lowest thresholds at 500, 1000, and 2000 Hz), whereas the 60% points for the 5- and 3-year-olds were 5 and 9 dB SL, respectively. The TIP, DIP, and NU-CHIPS are closed-set, identification tasks, so the performances were better than would be expected on more open-set, recognition tasks (Wilson & Antablin, 1982). Finally in the quiet domain, Hnath-Chisolm, Laipply, and Boothroyd (1998) used a nonsense syllable, three-interval oddity paradigm with 5- to 11-year-old children and reported that the major age effect was observed in children less than 7 years of age.

In the speech-in-noise domain, Elliott (1979), who administered the Speech Perception in Noise Test (SPIN; Kalikow, Stevens, & Elliott, 1977), reported two findings: (a) performances by 17-year-olds approximated performances reported for young adults by Kalikow et al. (1977) and (b) performances by 15- and 17-year-old children were better than performances by 11- and 13-year-old children. The differences, however, were significant only for the high-predictability words (sentences) presented at 0 dB S/N. Elliott suggested that the younger children were not as efficient as the older children in using the "pointer words" provided in the high-predictability SPIN sentences (i.e., the top-down processing by the older children was more effective than the top-down processing by the younger children). Stuart et al. (2006) used the NU-CHIPS (Chermak, Pederson, & Bendel, 1984; Elliott & Katz, 1980) presented in continuous and interrupted noises at three SNRs to sixteen 4- to 5-year-olds and eight young adults. In quiet, identification performance by the children was about 7% poorer than performance by the adults. In noise, the identification–performance differences between the groups of listeners increased to 15 to 18 dB. More recently, Bradley and Sato (2008) used the Word Identification by Picture Identification (WIPI; Ross & Lerman, 1970), presented in classroom settings at multiple SNRs, to study performances by 6-, 8-, and 11-year-old children (n = 840). At the 60% point on their functions (their Figure 2, p. 2080), the 11-year-old children performed about 10 dB better than the 6-year-old children, a difference that was reduced to 7 dB at the 95% point on the functions (their Figure 4, p. 2081).

Three studies have investigated the effect of children's age on the recognition performance on NU No. 6. In the first study, Larson, Petersen, and Jacquot (1974) compared recognition performances by forty 5- and 6-year-old children and forty 20- to 26-year-old adults. The recorded NU No. 6 materials were presented in a sound field in quiet and in white noise at 10-, 15-, and 20-dB S/N. In quiet, the adults performed 2.2% better than the children, but as the SNR decreased, the differences in performances increased from 12.2% at 20-dB S/N to 19.7% at 10-dB S/N. The Larson et al. study provides a good demonstration of the detrimental effects that noise has on the speech perception performance of young children as compared with adults. The second study, conducted by Chermak and Dengerink (1981), examined recognition performance on the NU No. 6 (Auditec recording) in quiet (40 dB SL) and in broadband noise (0 dB S/N) by 7-, 9-, 11-, and 13-year-old children and adults. Performances by all groups in quiet ranged from

[1]The use of *identification* implies a response to a closed set of materials, such as a picture-pointing task, whereas *recognition* implies a response in a more open-set paradigm (Wilson & Margolis, 1983).

95.0% to 99.7%, whereas performances in noise ranged from 25.7% to 30.7%. There were no significant differences, either for age or for Age × Condition. The lack of differences in quiet is attributable to the high presentation level of the words, and in noise it is difficult to explain. The third study, conducted by Sanderson-Leepa and Rintelmann (1976), involved comparisons with the WIPI, the Phonetically Balanced Kindergarten Test (PBK; Haskins, 1949), and the NU No. 6, all presented in quiet sound fields to groups of 3-, 5-, 7-, 9-, and 11-year-olds. All materials were recorded by the same speaker. The best performance was on the WIPI; it was attributed to the WIPI being an identification task (closed-set response) as compared with the recognition task involving the PBK items and NU No. 6 (open-set response). In addition, the PBK items were easier than the NU No. 6 items. This finding, however, was based on incomplete psychometric functions for both materials, with the majority of datum points >80% correct.

These three reports provide a sampling of the multitude of studies that consistently demonstrate that as children mature from their younger years into their teenage years, performances on auditory listening tasks improve. This dynamic behavior as a function of age necessitates the establishment of age-related norms for any behavioral speech intelligibility task. Thus, as previously mentioned, the purpose of the present study was the development of WIN age-related norms for 6- to 12-year-old children.

Method

Participants

Forty-two children in each of seven age groups (6–12 years) participated along with 24 young adults (18–27 years). The mean ages and standard deviations are listed in Table 1. Of the 294 children, 49% were females, 51% were males, 92% were Caucasian, and 8% were non-Caucasian; 12 of the young adults were males. In accordance with the East Tennessee State University Institutional Review Board regulations, written parental consent was obtained for each of the children, and informed consent was obtained for each of the adults. For inclusion, the participants were required to have normal otoscopy, to pass a pure-tone screening (500, 1000, 2000, and 4000 Hz) at 15 dB HL (American National Standards Institute, 2004) in the test ear, and to have no apparent physical or mental condition that would limit participation in the project (e.g., a phonological deficit).

Materials

As previously described, the clinical version of the WIN consists of two 35-word lists (female speaker; Department of Veterans Affairs, 2006). For this study, two randomizations of each of the two lists were copied digitally onto a CD along with a list of the 70 NU No. 6 words in quiet that are included in the WIN. A 1000-Hz tone was included for calibration.

Procedure

After the pure-tone screening, the three 35-word WIN lists (trials) were administered to the test ear using one of the following two sequences that were alternated across the participants: (a) List 1, List 2, List 1 or (b) List 2, List 1, List 2. The three lists that were presented to each participant can be thought of as three trials, designated *WIN-1, WIN-2,* and *WIN-3*. WIN-1 and WIN-3 were different randomizations of one of the two WIN lists, which provided a test–retest measure. WIN-2 was a randomization

Table 1. The means and standard deviations for the ages (years), Northwestern University Auditory Test No. 6 (70 NU No. 6) words in quiet (percentages), and the adjusted Peabody Picture Vocabulary Test—Revised (PPVT–R) scores for the seven groups of 42 children each.

	Group (years)						
Variable	6	7	8	9	10	11	12
Age							
M	6.6	7.5	8.4	9.6	10.5	11.5	12.5
SD	0.3	0.3	0.3	0.3	0.3	0.3	0.3
NU No. 6 in quiet							
M	93.1	96.8	98.0	98.5	98.5	98.3	98.8
SD subjects	6.1	2.3	1.8	1.6	1.8	1.7	1.1
SD words	6.2	3.9	4.6	2.6	2.7	3.4	3.3
PPVT–R (adjusted)							
M	105.2	106.4	110.1	110.5	110.2	112.3	107.8
SD	10.4	13.9	13.5	13.8	14.0	14.7	13.6

of the WIN list that was not used in WIN-1 and WIN-3. The combinations of two WIN lists and two randomizations of each list provided 8 three-list sequences that were used consecutively throughout the 318 participants. In this manner, each of the two lists was used an equal number of times in each of the three trials. The noise was presented at 70 dB SPL with the words presented at seven SNRs from 24 to 0 dB in 4-dB decrements (94–70 dB SPL). After the WIN lists, the 70 NU No. 6 words that comprise the WIN were presented in quiet at 90 dB SPL, which corresponded to the word presentation level during the 20-dB S/N condition. This was done to ensure that the children could repeat the WIN words under a favorable listening condition. Finally, as a descriptive reference, the Peabody Picture Vocabulary Test—Revised, Form L (PPVT–R; Dunn & Dunn, 1981) was administered. The young adults completed only the WIN portion of the experiment. The WIN materials were reproduced by a CD player (Sony Model CDP-CE375) and fed through an audiometer (Maico Model MA-41) to a TDH-50 earphone in an MX-41/AR cushion. Testing was monaural, with the right ears used on even-numbered subjects and left ears used on odd-numbered subjects (i.e., an equal number of right and left ears were examined in each age group). The participants verbally repeated the WIN words with the responses recorded into a spreadsheet. Testing took about 15 to 20 min per child.

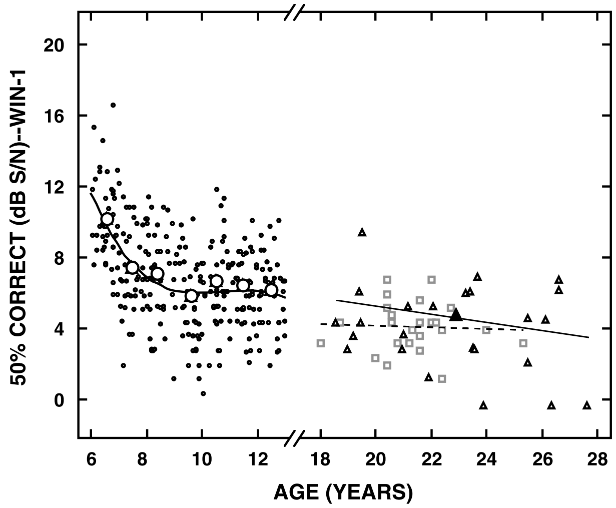
All testing took place in a relatively quiet room at each of the six participating institutions (see Acknowledgments). The noise levels in the test rooms were <50 dBA (Radio Shack Model 33-2055) and quiet enough to complete the hearing screening at 15 dB HL. Because the normative data for the original WIN were obtained from young adults with normal pure-tone thresholds tested in a sound booth, 24 young adults were included in this study to determine whether the test environment (i.e., a quiet room vs. a sound booth) influenced subject performance.

Results

WIN Test

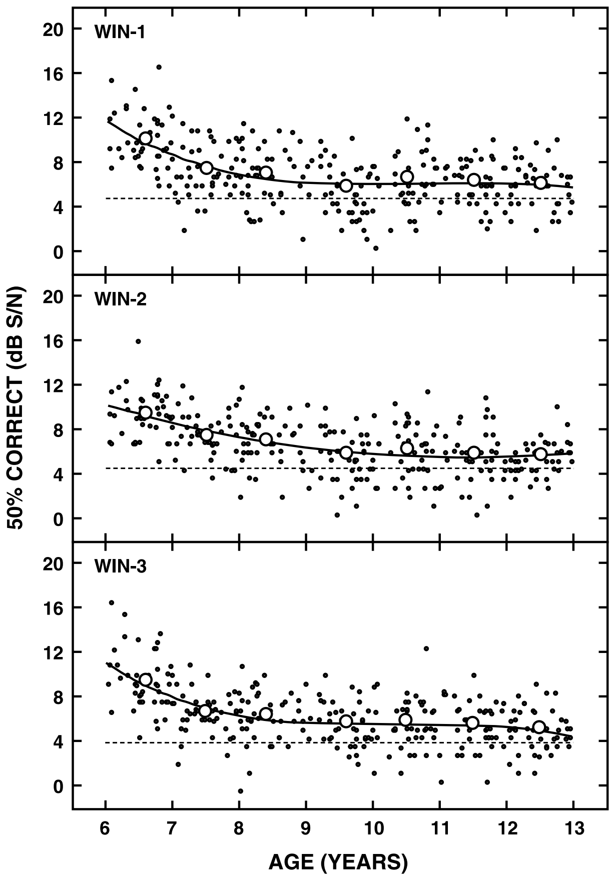
The WIN data from the individual listeners were quantified in terms of the 50% correct point calculated with the Spearman–Kärber equation (the equation is detailed in the Appendix.) The percentage correct at each of the seven SNRs also was evaluated for each listener/condition. The general results are illustrated in Figure 1, in which the 50% points (in dB SNR) for the 294 children from WIN-1 are plotted as a function of age (in years) along with data from two groups of 24 young adults with normal hearing for pure tones. One group was the 24 young adults from the Wilson et al. (2003) study, on whom the original WIN norms were established (gray squares); the second group was the 24 young adults included in the present study (triangles). The data in the figure were jittered.[2] Several features in Figure 2 deserve mention. First, the biggest change in recognition performance occurs between the ages of 6 and 7 years. Second, from 9 to 12 years, performance is stable. Third, performance by young adults (18–27 years) is slightly better (~1.5 dB) than performance by the older children. We examined these observations for WIN-1 using a one-way analysis of variance (ANOVA), which revealed a main effect of group, $F(7, 317) = 23.38$, $p < .001$, with a partial η^2 of .49. Post hoc analyses using multiple t tests with Bonferroni corrections showed that the 6-year-old group was significantly different from all of the other age groups. Similarly, the oldest group (young adults) also was significantly different from all of the other age groups. The only other statistically significant difference among all the age groups on WIN-1 was between the 7- and 9-year-old groups. Finally, the mean performance by the young adults (M = 4.8-dB S/N, SD = 2.2) is similar to the mean performance by the young adults in Wilson et al.'s (2003) study (M = 4.1-dB S/N, SD = 1.4), which is less than a one-word difference.

Figure 1. The first Words in Noise Test trial (WIN-1) 50% correct points (jittered) derived with the Spearman–Kärber equation from the 294 children (filled circles) and the 24 young adults (triangles) shown as a function of age. The large open circles are the means for the seven age groups, and the line is the best fit, third-degree polynomial. A linear regression (solid line) is used to describe the data from the young adults. The squares and dashed line represent the data from another sample of young adults used to norm the WIN (Wilson et al., 2003). S/N = signal-to-noise ratio.



[2] To avoid superimposed datum points, the data were jittered in several of the figures using a randomized, multiplicative algorithm between 1.025 and 0.975 in 0.005 steps. This accounts for the instances in which scores appear to exceed 100%.

Figure 2. The 50% correct points (jittered) derived with the Spearman–Kärber equation from the 294 children (filled circles) shown as a function of age for the three WIN tests (trials) in the protocol. The large open circles are the means for the seven age groups, and the line is the best fit, third-degree polynomial. The dashed line represents the mean data from the 24 young adults for the respective WIN trials.



A comparison of recognition performances on WIN List 1 and List 2 was possible because each of the two WIN lists was administered an equal number of times in WIN-1 and WIN-2. For the 294 children, the overall means were 7.2- and 6.9-dB S/N for List 1 and List 2, respectively, with corresponding standard deviations of 2.5 and 2.3 dB. A paired-sample t test revealed no significant differences between WIN List 1 and List 2 ($p > .05$). In addition, paired-sample t tests for each age group showed no significant list difference within each age group. Given that no list difference was found, we collapsed the data across lists to examine the effect of trial.

The individual 50% points calculated with the Spearman–Kärber equation for the three trials (WIN-1, WIN-2, and WIN-3) are plotted in Figure 2 as a function of the age of the listeners. Again the data were jittered. The means for each age group are depicted by the large open circles, with third-degree polynomials used to describe the data. As a point of reference, the 50% correct points of the 18- to 27-year-olds on each of the three trials are depicted by the dashed line. The means and standard deviations are listed in Table 2. The data for WIN-2 and WIN-3 closely mirror the data from WIN-1, with the main difference being slightly improved performance across the three trials. Within each of the three trials, the relations across age remained constant, with the greatest changes occurring between the 6- and 7-year-old groups and continued improvements between the 7- and 8-year-old groups and between the 8- and 9-year-old groups. Mean performance was fairly consistent among the upper four age groups of children (9–12 years), with only about a 1.5-dB difference between the mean performances by the 12-year-olds and the young adults.

In Figure 3, the individual 50% point data are recast into bivariate plots of WIN-1 versus WIN-2 (top panel) and WIN-1 versus WIN-3 (bottom panel). The diagonal line in each panel represents equal performance, with points above the line representing better performance on WIN-1 and points below the line representing better performance on the WIN-2 or WIN-3. The data in Figures 2 and 3 and Table 2 reflect that the relations described in Figure 1 were maintained across the three trials of the WIN. To examine the effect of trial, we used a mixed-model, repeated measures version of the general linear model ANOVA to analyze the data. The within-subject variable was trial (WIN-1, WIN-2, and WIN-3), whereas the between-subjects variable was age group. There was a main effect of age group that was consistent with the results of the previously described one-way ANOVA. In addition, a main effect of trial was found to be significant, $F(2, 620) = 12.16$, $p < .001$; however, there was no significant interaction between trial and age group. The partial η^2 was .04. Post hoc analyses using multiple t tests with Bonferroni corrections showed no significant difference between Trial 1 (M = 6.7-dB S/N) and Trial 2 (M = 6.5-dB S/N); however, Trial 3 (M = 6.2-dB S/N) was significantly different from both Trials 1 and 2. On average, performance at the 50% point improved 0.2 dB between WIN-1 and WIN-2 and 0.3 dB between WIN-2 and WIN-3. When one considers that each token in the Spearman–Kärber equation has a value of 0.8 dB (i.e., 4-dB step size divided by five words/step; see Appendix), one realizes that the 0.5- and 0.3-dB differences between WIN-3 and WIN-1 and between WIN-3 and WIN-2, although statistically significant, are small. These effects are apparent in the individual data presented in Figure 3. In both panels of Figure 3, better performance was achieved by more children on WIN-2 than on WIN-1 (48.3% vs. 39.1%) and on WIN-3 than on WIN-1 (57.8% vs. 41.2%). Because the WIN lists used with WIN-1 and WIN-2 were different and were presented in counterbalanced fashion, the small improvement in performance

Table 2. The means, standard deviations, and 90th percentiles (dB signal-to-noise [S/N]) for each of the three Words-in-Noise Test (WIN) trials based on the individual data from the Spearman–Kärber equation.

Measure	Group (years)							
	6	7	8	9	10	11	12	18–27
WIN-1								
M	10.2	7.5	7.1	5.9	6.7	6.5	6.2	4.8
SD	2.4	2.2	2.4	2.1	2.2	1.7	1.8	1.9
90th	13.2	10.7	10.0	8.4	9.2	8.4	8.4	6.6
Poly 50%	8.0	6.0	5.9	4.5	5.5	5.1	5.3	3.9
Slope at 50%	6.4	6.8	6.4	7.3	6.9	7.2	7.1	7.4
WIN-2								
M	9.6	7.6	7.2	5.9	6.4	6.0	5.8	4.6
SD	2.0	1.8	2.0	2.0	2.1	1.9	1.3	1.4
90th	12.3	9.9	9.9	8.3	9.1	9.1	6.8	6.6
Poly 50%	8.2	6.4	6.4	4.4	5.3	4.8	5.1	3.8
Slope at 50%	6.3	6.7	6.6	7.4	7.0	6.9	6.7	6.7
WIN-3								
M	9.6	6.9	6.6	5.9	6.1	5.8	5.4	4.0
SD	2.6	1.6	1.9	1.8	2.0	1.9	1.6	1.7
90th	13.2	9.1	9.1	8.3	8.3	8.4	7.5	5.8
Poly 50%	8.0	5.6	5.4	4.5	4.7	4.8	4.3	2.8
Slope at 50%	6.1	7.3	6.6	7.6	7.4	7.0	7.3	7.5
Suggested cutoff	13.2	10.0	10.0	8.4	8.4	8.4	8.4	6.8
No. correct	16	20	20	22	22	22	22	24

Note. Also listed are the 50% points (dB S/N) and the slopes at the 50% points (%/dB) calculated from the polynomials (and first derivatives of the polynomials) used to describe the data depicted in Figure 3. The 90th percentiles listed at the bottom of the table are the cutoffs suggested for the respective age groups; "No. correct" corresponds to the suggested cutoffs. WIN-1 and WIN-3 = different randomizations of one of the two WIN lists; WIN-2 = a randomization of the WIN list that was not used in WIN-1 and WIN-3; Poly 50% = 50% points calculated from the polynomials used to describe the data.

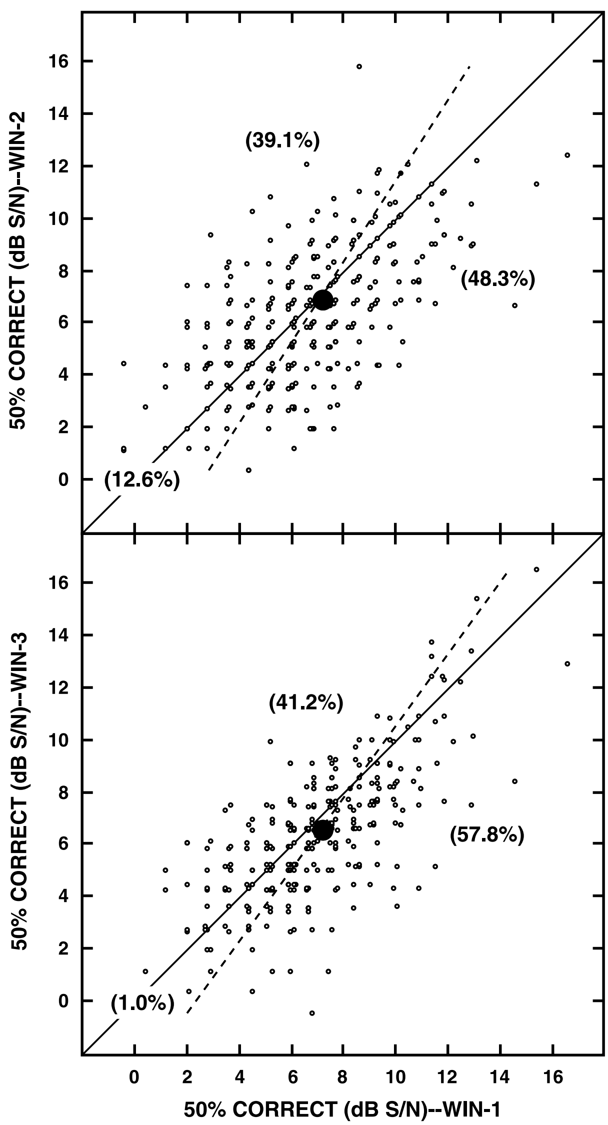
suggests the listeners were learning to listen to the words in the multitalker babble background and were not learning the test materials. The same learning probably continued in WIN-3, but the comparison between WIN-1 and WIN-3 was compounded by repetition of the same WIN list (different randomization) that might have produced some minimal learning effect of the stimulus words. In any event, the change in performance across the three trials was small (<1 dB). The Pearson product–moment correlation between WIN-1 and WIN-2, collapsed across age groups, was .62 ($p < .001$), whereas that between WIN-1 and WIN-3 was .74 ($p < .001$). These correlations are consistent with the fact that WIN-1 and WIN-2 are similar but different lists and WIN-1 and WIN-3 are the same lists. The high correlation between WIN-1 and WIN-3 indicates good test–retest reliability.

To complete the description of the results, the psychometric functions for each of the three WIN trials are depicted for each of the age groups in Figure 4. The morphology of the functions in the figure is strikingly similar, both across trials in a given age group and across age groups. Only the locations of the functions with respect to the x-axis change noticeably across groups. The 50% points calculated from the polynomials used to describe the data are listed in Table 2 (Poly 50%) along with the slopes of the functions at the 50% points. As with the previous WIN data (Wilson, 2003; Wilson, Carnell, & Cleghorn, 2007), the 50% points calculated from the mean psychometric function are consistently 1 to 2 dB lower SNRs than are the 50% points calculated with the Spearman–Kärber equation. The reason for this disparity rests with the Spearman–Kärber equation, in which performance is assumed to extend systematically from 0 to 100% but because of noise in the data does not occur in all instances.

NU No. 6 in Quiet

The mean percentage correct recognitions for the 70 NU No. 6 words used in the WIN are listed by age in Table 1. Recall that these 70 words were presented in quiet at 70 dB SPL. Both the intersubject and interword standard deviations are included. The mean was lowest for the 6-year-old group and asymptotic for the 8- to

Figure 3. Bivariate plots of the individual 50% points for WIN-1 and WIN-2 (top panel) and WIN-1 and WIN-3 (bottom panel). To help avoid overlapping datum points, the data were jittered. The large circles represent the mean data. The dashed lines are the linear regressions used to describe the data. The numbers in parentheses are the percentages of datum points above, on, and below the diagonal line that represents equal performance.



12-year-old groups. In the 6-year-old group, one child scored 70% correct, seven scored between 80.0% and 86.3% correct, and the rest scored >90% correct. All children in the remaining six age groups scored >90% correct. For all groups, two words in quiet were correct <90%: (a) *mess* (82.3%) and (b) *calm* (84.0%). In the WIN Test, these words are presented at 4 and 0 dB S/N, respectively, and it is doubtful they played a substantial role in the WIN results. The individual recognition scores on the 70 NU No. 6 words are plotted in Figure 5 as a function of the 50% correct points (dB S/N) obtained on WIN-1. The data in the figure indicate that the just-mentioned eight 6-year-olds who scored <90% correct on the words presented in quiet also had poorer performances on the WIN. These data in quiet suggest that NU No. 6 can be used efficaciously with children as young as 6 years. For the statistical analyses, we transformed the data with rationalized arcsine units (Studebaker, 1985). A one-way ANOVA showed a significant between-groups effect on speech recognition in quiet, $F(6, 293) = 12.7$, $p < .001$. Post hoc analyses using a Bonferroni correction for multiple t tests showed that the 6-year-olds were significantly different from all other age groups. In addition, the 7-year-olds were significantly different from the 12-year-olds. All other comparisons among age groups were not statistically different. Although the differences among some of the groups were statistically significant, the mean differences were less than three tokens on a 50-word list (6%) and are not considered of clinical importance.

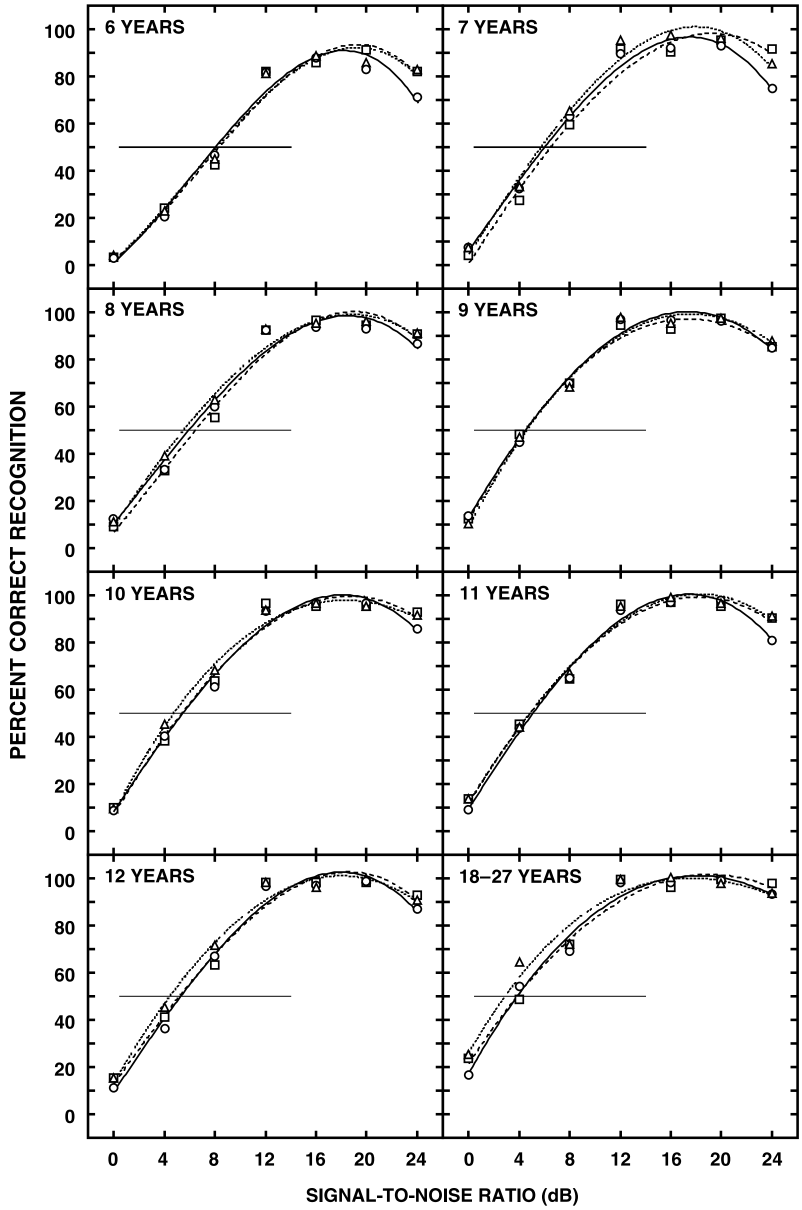
PPVT–R

The mean adjusted scores on the PPVT–R and standard deviations obtained for the seven age groups are listed in Table 1. The adjusted mean scores ranged from 105.2 (6-year-olds) to 112.3 (11-year-olds) and were all within the normal 85–115 range. Nine children scored below 85 on the PPVT–R. Each of these nine children had >90% correct word recognition performances on the NU No. 6 in quiet. Likewise, as the data in Figure 6 show, below-normal performances by these nine children on the PPVT–R did not translate to abnormal performances on the WIN. Figure 6 is a bivariate plot of the PPVT–R scores and the 50% correct points (dB S/N) on the WIN-1. The data exhibit a random pattern that is supported by the slope of the linear regression used to describe the data, which approaches zero. The data in the figure indicate the lack of a relationship between performance on the PPVT–R and performance on the WIN, at least within the range of scores involved. The lack of relationship was confirmed by an insignificant Pearson product–moment correlation (–.07) between WIN-1 and PPVT–R score.

Discussion

The effect of age on recognition performance with the WIN was most dramatic between 6 and 7 years of age, when mean performance on WIN-1, WIN-2, and WIN-3 changed 2.5 dB, from 9.8-dB S/N (6 years) to 7.3-dB S/N (7 years). All other changes between adjacent age groups of the children were 1 dB or less. This finding of the largest change in WIN performance between 6 and 7 years is in agreement with the majority of the literature, exemplified by the Hnath-Chisolm et al. (1998) study, in which

Figure 4. Percentage correct at the seven S/N ratios on the three WIN trials (WIN-1 = circles and solid lines, WIN-2 = squares and dashed lines, WIN-3 = triangles and dotted lines) for the eight groups of listeners. The lines connecting the datum points are the best fit, third-degree polynomials that had R^2 values between .97 and .99.



age was a significant factor in performance for children below age 7. The present data further suggest that recognition performance was stable between 9 and 12 years but still slightly poorer than the performance achieved by young adults over age 17. This age factor also was reported by Elliott (1979), who stated that performances on the SPIN by 17-year-olds were similar to performances by young adults but that performances by 15- and 17-year-olds were better than performances by 11- and 13-year-olds.

Regarding the poorer performance by the 6-year-olds on the WIN, the question is "Was that 2.5- to 3.9-dB poorer performance (see Table 2) owing to a lack of knowledge about or familiarity with the target words or to a reduced ability to segregate the target words from the background noise?" Recognition performance on the NU No. 6 words presented in quiet suggests that both factors contribute to the poorer performance by the 6-year-olds. In quiet, the 6-year-olds performed 3.7% to 5.7% more poorly than the other age groups. Considering the slope

Figure 5. Percentage correct on the 70 Northwestern University Auditory Test No. 6 words in quiet (ordinate) for the 294 listeners versus the 50% correct points (dB S/N) on WIN-1. For graphic clarity, the data were jittered. The line is a best fit, second-degree polynomial used to describe the data.

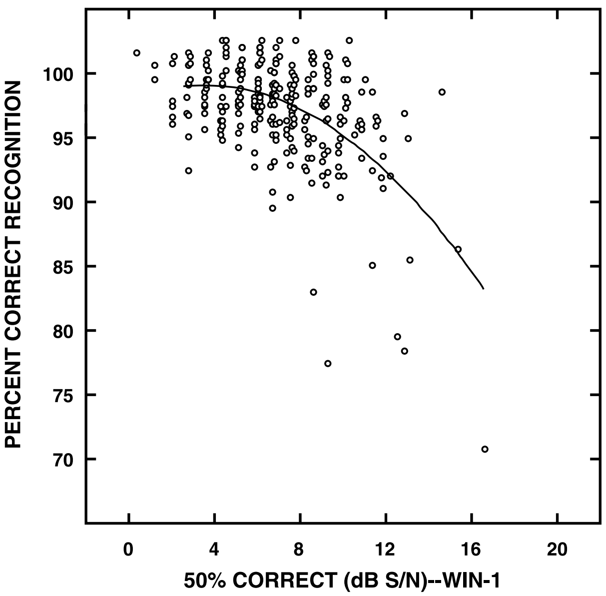
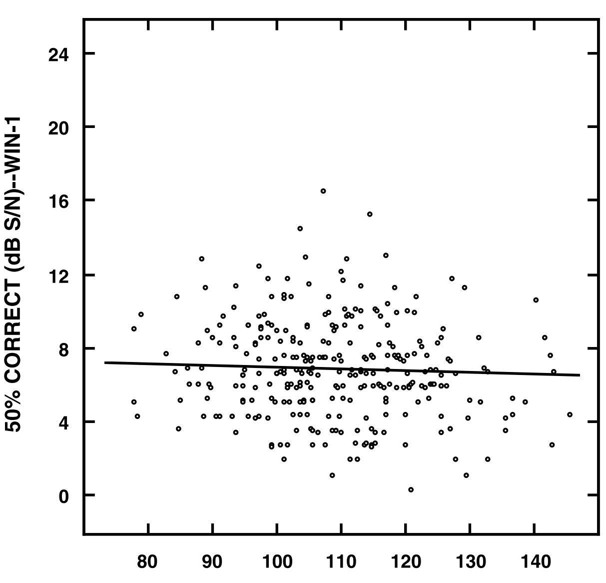


Figure 6. The 50% correct points (dB S/N) for the 294 listeners plotted as a function of their adjusted score on the Peabody Picture Vocabulary Test—Revised. The linear regression used to describe the data has a slope of –0.0087 dB/point.



of the psychometric functions, these percentage differences could account for about 1 dB of the 2.5- to 3.9-dB differences that were observed between age groups on the WIN. The remaining 1.5- to 2.9-dB differences observed on the WIN must be attributable to an inability on the part of the children to segregate the words from the multitalker babble. This inability could be related to maturation in the form of listening strategy issues, neurophysiologic issues, or a combination of both.

Normative data on the original WIN were determined from 24 young adult listeners with normal pure-tone thresholds using the 90th percentile, which was a 6-dB S/N, as the upper cutoff of normal performance (Wilson et al., 2003). The data in Figure 5 indicate that 118 (40.1%) of the 294 children performed in the normal range for adults on WIN-1 (i.e., ≤6-dB S/N). Normal performances on WIN-1 ranged from two children in the 6-year-old group (4.8%) to 25 children in the 9-year-old group (59.5%). Over the course of the three trials, normal performances on the WIN increased from 40.1% (WIN-1) to 49.0% (WIN-3). Because the performances by the children with normal hearing were at higher SNRs than the performances by the young adults on whom the WIN was standardized, different sets of upper cutoff values were required. We calculated the 90th percentiles for each of the three WIN trials by each of the age groups. These 90th percentiles are listed in Table 2. A suggested upper cutoff based on these three values for each age group was determined for each age group and is listed at the bottom of the table. These suggested cutoffs are rounded to the nearest actual value (decibel SNR) that is achievable with the Spearman–Kärber equation, which is the metric we used to evaluate the individual performances on the WIN. For example, the 12.3-dB S/N cutoff calculated for the 6-year-olds' WIN-2 data corresponds to 17.125 words correct, which is not possible on an individual listener basis. Because two of the three 90th percentiles with the 6-year-olds were 13.2-dB S/N, we selected that value, which corresponds to 16 correct responses, as the normal upper cutoff for the 6-year-olds. We used similar processes to establish the suggested upper cutoffs for the remaining age groups.

The data from this study indicate that both the WIN and the NU No. 6 materials presented in quiet can be used with children as young as 6 years of age. Even with children as old as 12, the WIN requires that the ranges of normal performance require adjustment upward in SNR compared with the norm used with adults. Suggested upper cutoffs for each of the seven age groups ranged from 13.2 dB S/N (6-year-olds) to 8.4 dB S/N (9- to 12-year-olds). The mean recognition performances on the 70 NU No. 6 words presented in quiet, which ranged from 93.1% (6-year-olds) to 98.8% (12-year-olds), were interpreted as an indication that the materials were appropriate for use with children in this age range.

Acknowledgments

Support for this study was provided by the East Tennessee State University Foundation. The Rehabilitation Research and Development Service, Department of Veterans Affairs, supported this work through a Senior Research Career Scientist award to the first author. This study was made possible by the following institutions: Ashley Academy, Boys and Girls Club, St. Mary's School, and University School (Johnson City, TN); Tri-Cities Christian School (Blountville, TN); and Flatwoods Elementary (Jonesville, VA). Appreciation is expressed to Rachel McArdle for her expertise with inferential statistics. The contents of this article do not represent the views of the Department of Veterans Affairs or the U.S. government. Portions of this article were presented at the American Speech-Language-Hearing Association Annual Convention in New Orleans, LA (2009); at the annual meeting of the American Auditory Society in Scottsdale, AZ (2010); and at the American Academy of Audiology Convention in San Diego, CA (2010).

References

American National Standards Institute. (2004). *Specification for audiometers* (ANSI S3.6-2004). New York, NY: Author.

Bradley, J. S., & Sato, H. (2008). The intelligibility of speech in elementary school classrooms. *The Journal of the Acoustical Society of America, 123,* 2078–2086.

Chermak, G. D., & Dengerink, J. (1981). Word identification in quiet and in noise. A re-examination. *Scandinavian Audiology, 10,* 55–60.

Chermak, G. D., Pederson, C. M., & Bendel, R. B. (1984). Equivalent forms and split–half reliability of the NU-CHIPS administered in noise. *Journal of Speech and Hearing Disorders, 49,* 196–201.

Dawes, P., & Bishop, D. V. M. (2008). Maturation of visual and auditory temporal processing in school-aged children. *Journal of Speech, Language, and Hearing Research, 51,* 1002–1015.

Department of Veterans Affairs. (2006). *Speech recognition and identification materials: Disc 4.0* [Computer software]. Mountain Home, TN: Veterans Affairs Medical Center.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test—Revised*. Circle Pines, MN: AGS.

Eisenberg, L. S., Shannon, R. V., Martinez, A. S., Wygonski, J., & Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *The Journal of the Acoustical Society of America, 107,* 2704–2710.

Elliott, L. L. (1979). Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. *The Journal of the Acoustical Society of America, 66,* 651–653.

Elliott, L. L., Connors, S., Kille, E., Levin, S., Ball, K., & Katz, D. (1979). Children's understanding of monosyllabic nouns in quiet and in noise. *The Journal of the Acoustical Society of America, 66,* 12–21.

Elliott, L. L., & Katz, D. R. (1980). *Development of a new children's test of speech discrimination*. St. Louis, MO: Auditec.

Fallon, M., Trehub, S. E., & Schneider, B. A. (2000). Children's perception of speech in multitalker babble. *The Journal of the Acoustical Society of America, 108,* 3023–3029.

Finney, D. J. (1952). *Statistical method in biological assay*. London, England: C. Griffen.

Goldman, R., Fristoe, M., & Woodcock, R. W. (1970). *Test of Auditory Discrimination*. Circle Pines, MN: AGS.

Haskins, H. (1949). *A phonetically balanced test of speech discrimination for children* (Unpublished master's thesis). Northwestern University, Evanston, IL.

Hnath-Chisolm, T. E., Laipply, E., & Boothroyd, A. (1998). Age-related changes on a children's test of sensory-level speech perception capacity. *Journal of Speech, Language, and Hearing Research, 41,* 94–106.

Jamieson, D. G., Kranjc, G., Yu, K., & Hodgetts, W. E. (2004). Speech intelligibility of young school-aged children in the presence of real-life classroom noise. *Journal of the American Academy of Audiology, 15,* 508–517.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America, 61,* 1337–1351.

Keogh, T., Kei, J., Driscoll, C., & Khan, A. (2010). Children with minimal conductive hearing impairment: Speech comprehension in noise. *Audiology and Neurotology, 15,* 27–35.

Larson, G. W., Petersen, B., & Jacquot, W. S. (1974). Use of Northwestern University Test No. 6 for speech discrimination testing with children. *Journal of Auditory Research, 14,* 287–292.

McCreery, R., Ito, R., Spratford, M., Lewis, D., Hoover, B., & Stelmachowicz, P. G. (2010). Performance-intensity functions for normal-hearing adults and children using Computer-Aided Speech Perception Assessment. *Ear and Hearing, 31,* 95–101.

Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *The Journal of the Acoustical Society of America, 87,* 2705–2715.

Palva, A., & Jokinen, K. (1975). Undistorted and filtered speech audiometry in children with normal hearing. *Acta Otolaryngologica, 80,* 383–388.

Ross, M., & Lerman, J. (1970). A picture identification test for hearing impaired children. *Journal of Speech and Hearing Research, 13,* 44–53.

Sanderson-Leepa, M. E., & Rintelmann, W. F. (1976). Articulation functions and test–retest performance of normal-hearing children on three speech discrimination tests: WIPI, PBK-50, and NU Auditory Test No. 6. *Journal of Speech and Hearing Disorders, 41,* 503–519.

Siegenthaler, B. M. (1969). Maturation of auditory abilities in children. *International Audiology, 8,* 59–71.

Siegenthaler, B. M. (1975). Reliability of the TIP and DIP speech-hearing tests for children. *Journal of Communication Disorders, 8,* 325–333.

Siegenthaler, B. M., & Haspiel, G. (1966). *Development of two standardized measures of hearing for speech by children* [Final report for U.S. Office of Education Cooperative Research Project 2372, Pennsylvania State University].

Washington, DC: U.S. Department of Health, Education, and Welfare.

Siegenthaler, B. M., Pearson, J., & Lezak, R. (1954). A speech reception threshold test for children. *Journal of Speech and Hearing Disorders, 19,* 360–366.

Stuart, A. (2005). Development of auditory temporal resolution in school-age children revealed by word recognition in continuous and interrupted noise. *Ear and Hearing, 26,* 78–88.

Stuart, A., Givens, G. D., Walker, L. J., & Elangovan, S. (2006). Auditory temporal resolution in normal-hearing preschool children revealed by word recognition in continuous and interrupted noise. *The Journal of the Acoustical Society of America, 119,* 1946–1949.

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *Journal of Speech and Hearing Research, 28,* 455–462.

Tillman, T. W., & Carhart, R. (1966). *An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University Auditory Test No. 6*. [U.S. Air Force School of Aerospace Medicine Technical Report]. Brooks Air Force Base, TX: U.S. Air Force.

Wilson, R. H. (2003). Development of a speech in multitalker babble paradigm to assess word-recognition performance. *Journal of the American Academy of Audiology, 14,* 453–470.

Wilson, R. H., Abrams, H. B., & Pillion, A. L. (2003). A word-recognition task in multitalker babble using a descending presentation mode from 24-dB S/N to 0-dB S/N. *Journal of Rehabilitation Research and Development, 40,* 321–328. Retrieved from http://www.vard.org/jour/03/40/4/pdf/Wilson-B.pdf.

Wilson, R. H., & Antablin, J. K. (1982). The Picture Identification Task—A reply to Dillon. *Journal of Speech and Hearing Disorders, 47,* 111–112.

Wilson, R. H., & Burks, C. A. (2005). The use of 35 words to evaluate hearing loss in terms of signal-to-noise ratio: A clinic protocol. *Journal of Rehabilitation Research and Development, 42,* 839–852. Retrieved from http://www.rehab.research.va.gov/jour/05/42/6/pdf/wilson.pdf.

Wilson, R. H., Carnell, C. S., & Cleghorn, A. L. (2007). The Words-in-Noise (WIN) Test with multitalker babble and speech-spectrum noise maskers. *Journal of the American Academy of Audiology, 18,* 522–530.

Wilson, R. H., & Margolis, R. H. (1983). Measurement of the auditory thresholds for speech stimuli. In D. Konkle & W. Rintelmann (Eds.), *Principles of speech audiometry* (pp. 79–126). Baltimore, MD: University Park Press.

Wilson, R. H., & McArdle, R. (2007). Intra- and inter-session test, retest reliability of the Words-in-Noise (WIN) Test. *Journal of the American Academy of Audiology, 18,* 813–825.

Wilson, R. H., McArdle, R., & Smith, S. L. (2007). An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. *Journal of Speech, Language, and Hearing Research, 50,* 844–856.

Wilson, R. H., Morgan, D. E., & Dirks, D. D. (1973). A proposed SRT procedure and its statistical precedent. *Journal of Speech and Hearing Disorders, 38,* 184–191.

Received December 9, 2009

Revision received February 2, 2010

Accepted March 25, 2010

DOI: 10.1044/1092-4388(2010/09-0270)

Contact author: Richard H. Wilson, Audiology (126/151), VA Medical Center, Mountain Home, TN 37684. E-mail: richard.wilson2@va.gov.

Appendix. The Spearman–Kärber equation.

We used the Spearman–Kärber equation (Finney, 1952; Wilson, Morgan, & Dirks, 1973) to calculate the 50% points from the data sets of each participant. The basic Spearman–Kärber formula is as follows:

$$50\% = i + \frac{1}{2}(d) - (d)(\#\ \text{correct})/(w),$$

in which i = the initial presentation level (dB S/N), d = the attenuation step size (decrement), and w = the number of items per decrement. For the Words in Noise (WIN) materials, which have 24 dB S/N as the highest presentation level and have five words for each 4-dB step, the equation was simplified to the following:

$$50\% = 26 - (\#\ \text{correct})(0.8).$$

The "0.8" is the attenuation step size (4 dB) divided by the number of words per step (5). The 50% point is expressed in dB S/N.